

Capstone Project-1

Play Store App Review Analysis

Team Name - kognetics

Rishabh Rohil
Saniya Bubere
Ashi Saxena
Rahul Jadhav

Why did We choose the google play store for analysis?

Google Play services helps to ensure the security and reliability of an Android device, and keep devices updated with the latest security features.

Google Play partners with developers to provide access to over 2 billion active monthly users and it offer nearly **2 Million** app. Distribution is free for everyone but because of this developers may receive feedback from tester and get certain amounts of reviews. We use this and got some knowledge about the app to predict its success

Mobile App Market is set to grow **20%** by **2023**

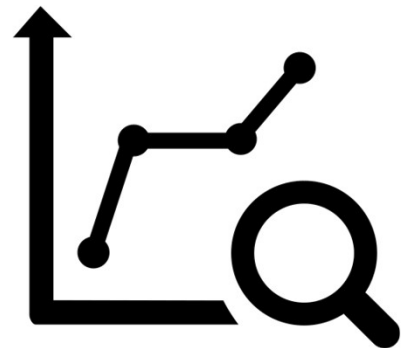
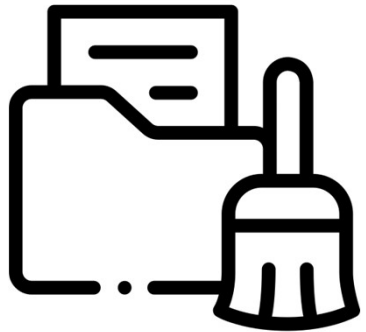
Android Apps comprise 90% of the Mobile App Market





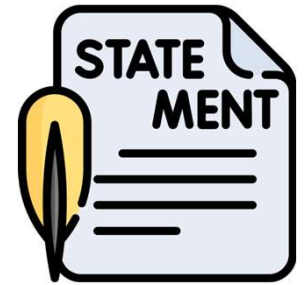
Overview of Exploratory Analysis

- ☐ **Explore the Data** : While Exploring the data we determine a useful way to understand what datasets are representing
- ☐ **Data Cleaning**: In Datasets there is so much Null values Because of that the prediction was not correct So We clean all the unwanted data from both datasets
- ☐ **Query Analysis**: In this We tried to make a visualization based on problems

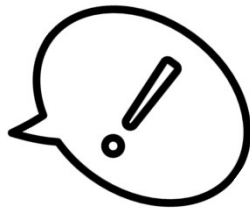




Problem Statement



- Two datasets are provided, one with basic information of Play Store App and Second one is of users reviews
- Every row contains various entries regarding a certain app. We will be doing Exploratory data analysis on this data set, which is a very important step in data science cycle, as it not only helps in taking very initial business decisions. Our objective will be to structure the data, clean it and present certain trends that we observe that can help us draw very preliminary conclusions about the probability of success of a newly launched app. Also our objective is to Explore and analyze the data to discover key factors responsible for app engagement and success
- We have to make a valuable insights from both the datasets in order to identify the important characteristics



Agenda:

- 1) Introduction
- 2) Sentiments of Free App
- 3) Sentiments of Paid App
- 4) Sentiments & Reviews
- 5) Correlation between Columns
- 6) box plots for size and rating
- 7) Percent of Free VS Paid App in Store
- 8) Number of Apps And its Category based on Type
- 9) Counts Based on Rating
- 10) Top 5 App Based on Category
- 11) Top 5 Category in terms of Installs
- 12) Ratings
- 13) Counts Based on Size
- 14) Average installs in term of Size
- 15) Count based on Android Version
- 16) Counts based on Last Updates
- 17) Updates with Years
- 18) Top 5 Revenue Categories

Attributes in Google Play store Data

Attributes	Meaning	Data Type
1.App	This column Contains the name of the app for each observation.	Qualitative
2.Category	This column Contains Category to which the app belongs.	Qualitative
3.Rating	This column contains the average rating for the app.	Quantitative
4.Reviews	This column contains the number of reviews that the app has received on the play store.	Quantitative
5.Size	This column contains the amount of memory the app occupies on the device.	Quantitative
6.Installs	This column contains the number of times that the app has been downloaded and installed from the play store.	Quantitative
7.Type	This column contains the information whether the app is free or paid.	Qualitative
8.Price	If the app is a paid app, this column contains the data about its price.	Quantitative
9.Content Rating	This column contains the maturity rating of the app i.e. the age group of the audience for which it is suitable.	Quantitative
10.Genres	This column contains the data about to which genre the app belongs. Genres can be considered as a further division of the group of Category.	Qualitative
11.Last Updated	Contains the date on which the latest update of the app was released.	Quantitative
12.Current Version	Contains information on the current version of the app available on the play store.	Quantitative
13.Android Version	Contains information about the android versions on which the app is supported.	Quantitative

Attributes in User reviews		
Attributes	Meaning	Data Type
1.App	Application name	Qualitative
2.Translated Review	User review	Qualitative
3.Sentiment	Positive/Negative/Neutral	Quantitative
4.Sentiment Polarity	Sentiment polarity score	Quantitative
5.Sentiment Subjectivity	Sentiment subjectivity score	Quantitative

Dataset Construction

- Mounting the drive : Because of Mounting we can access csv file to do analysis
- Loading Datasets : Two datasets is there in project First is store app and Second is User Reviews
- Import Libraries: Compulsory to use Libraries like Numpy , Pandas , Matplotlib and Seaborn .
- When your using Google colab then this necessary to use `%matplotlib inline` where plot graphics will appear in your notebook
- Data Cleaning: Removing Outliers , Finding and Removing Null values
- Data Assings: Replacing the Null Values with Mean ,Median & Mode with its suitable Condition.
- Query Analysis: Analayzing the data sets to points of interest using visualization.

Explore the Arrangement of Datasets

Play store dataset.csv: shape(10841,13)

This file contain 13 features that is:

- i. App: Name of Apps.
- ii. Category:Category of Apps (ART_AND_DESIGN,FINANCE,COMIC,COMMUNICATION,GAME etc)
- iii. Rating:Rating on app and it is between 0 to 5
- iv. Reviews:Total Users given reviews on every app
- v. Size:Size is in Megabytes(MB)
- vi. Installs:Number of times the app was downloaded from google play
- vii. Type:Type is Paid or Free
- viii. Price:Price is in \$
- ix. Content Rating:Which Age Group have Rated the App

Explore the Arrangement of User Reviews Datasets

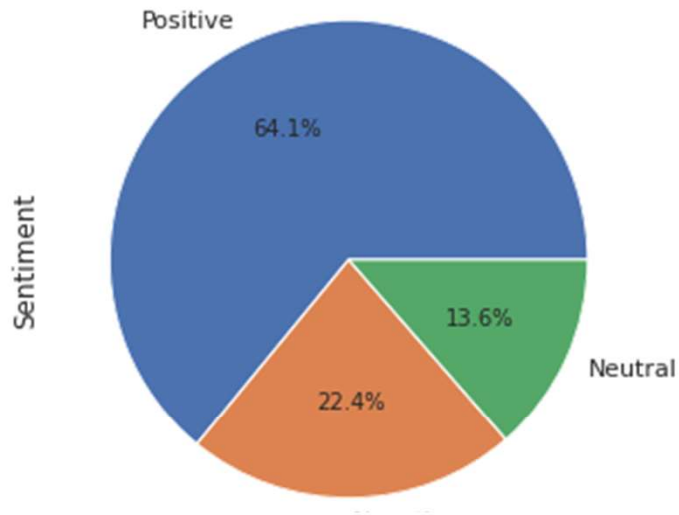
User reviews dataset.csv: shape(64295,5)

This file contain 5 features that is:

- i. App: Its contain Apps Name.
- ii. Translated Review: It contain reviews
- iii. Sentiments: It contain Positive ,Negative, Neutral Sentiments.
- iv. Sentiment Subjectivity:It contain Numeric values between negative to positive 1.
- v. Sentiment Polarity: defines the orientation of the expressed sentiment

Exploratory Data Analysis

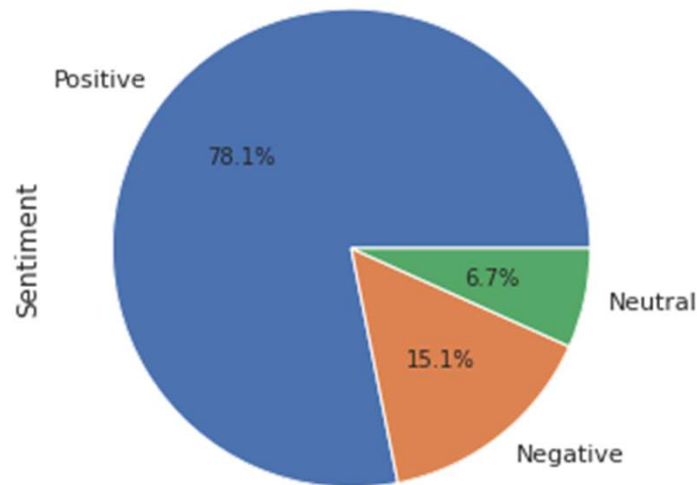
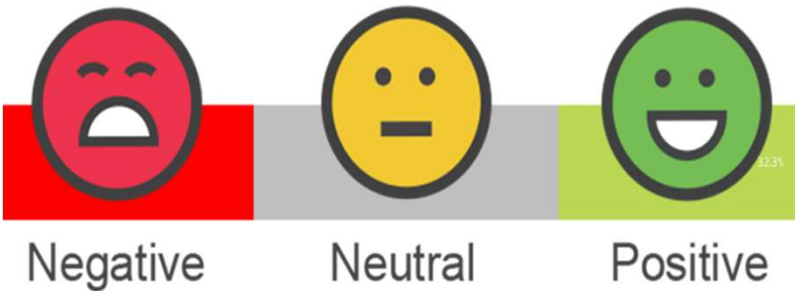
Sentiments of Free App



In the free apps 64.1 % apps has positive sentiment followed by 22.4 % and 13.6 % for Negative and Neutral sentiment respectively



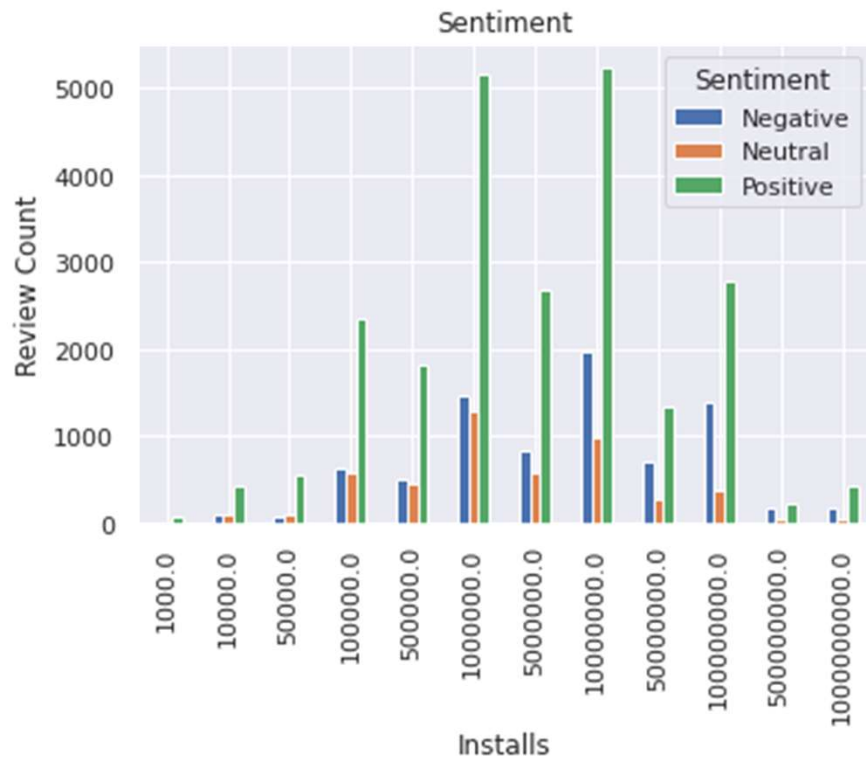
Sentiments of Paid App



In the paid apps 78.1 % apps has positive sentiment followed by 15.1 % and 6.7 % for Negative and Neutral sentiment respectively



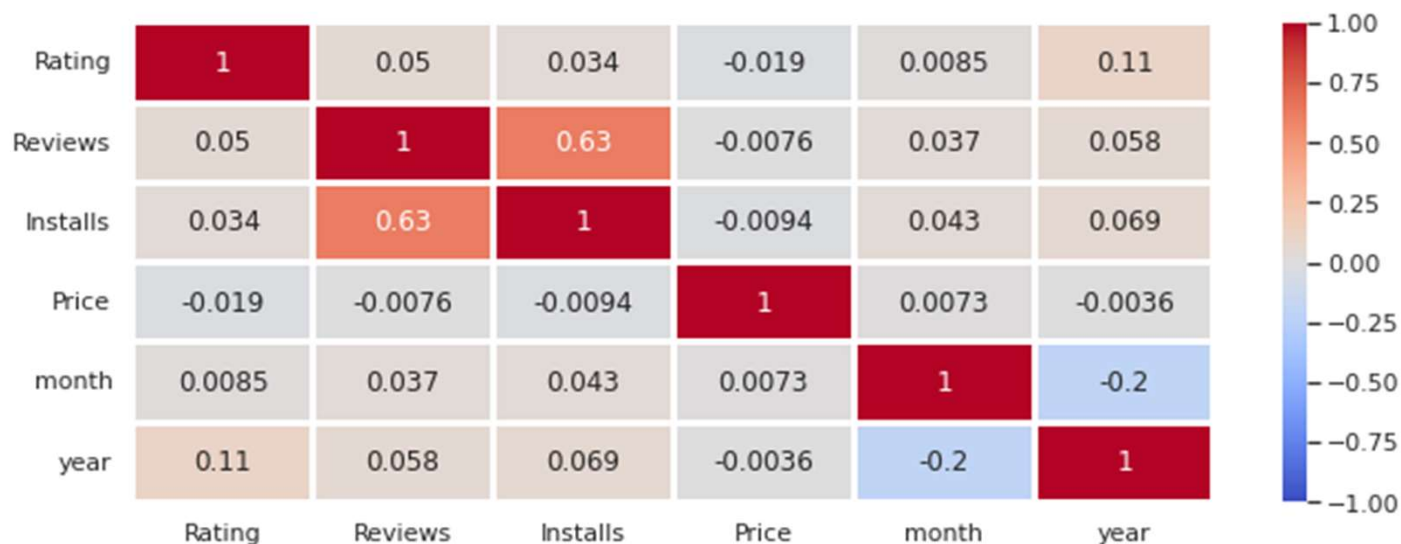
Sentiments And Reviews



From the Sentiment , review graph we can see that no of installs 1000000 and 10000000 has highest review with positive sentiments

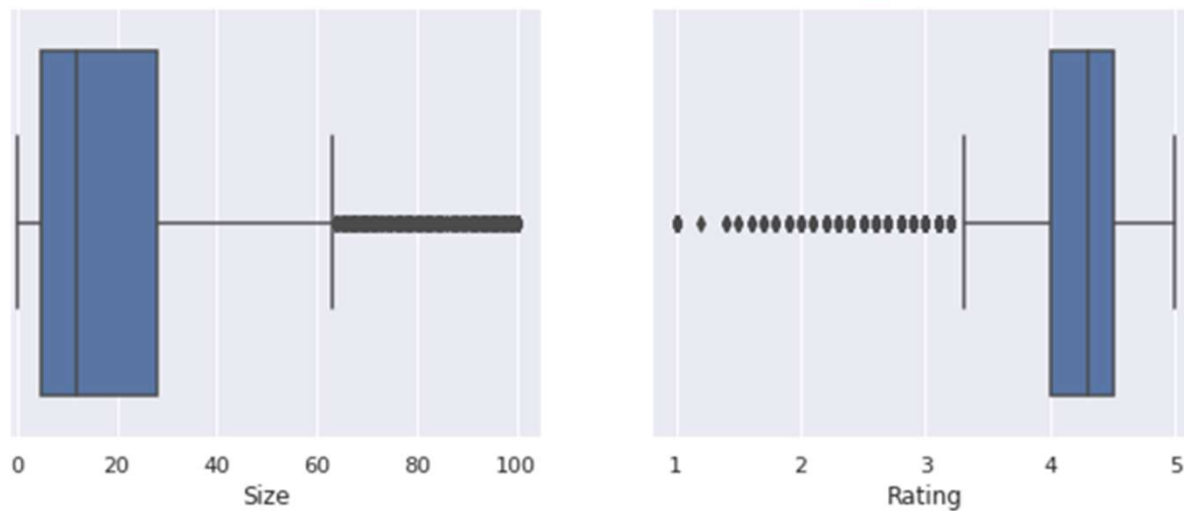


The correlation between different columns.



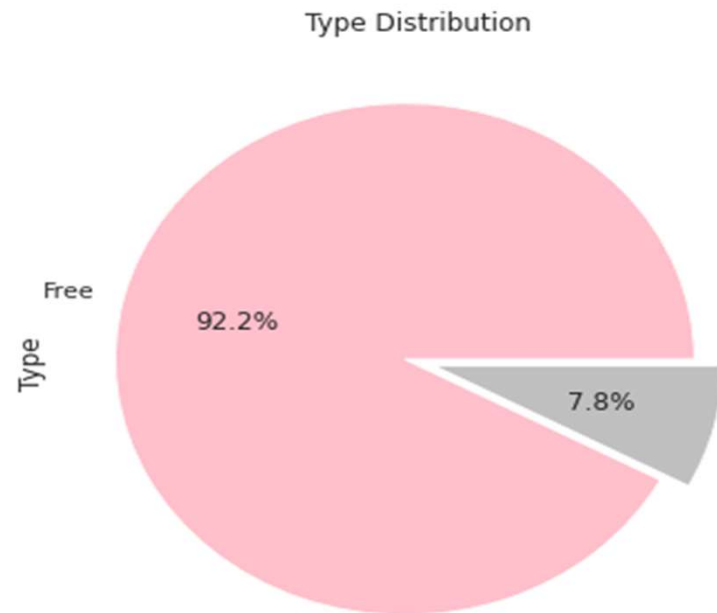
From this correlation matrix, we see that reviews and installs are positively correlated. Also, if an app has higher number of installations then obviously it will have more reviews. One more thing to notice is that correlation of price with every column is less than 0.

Box Plot to identify Outliers



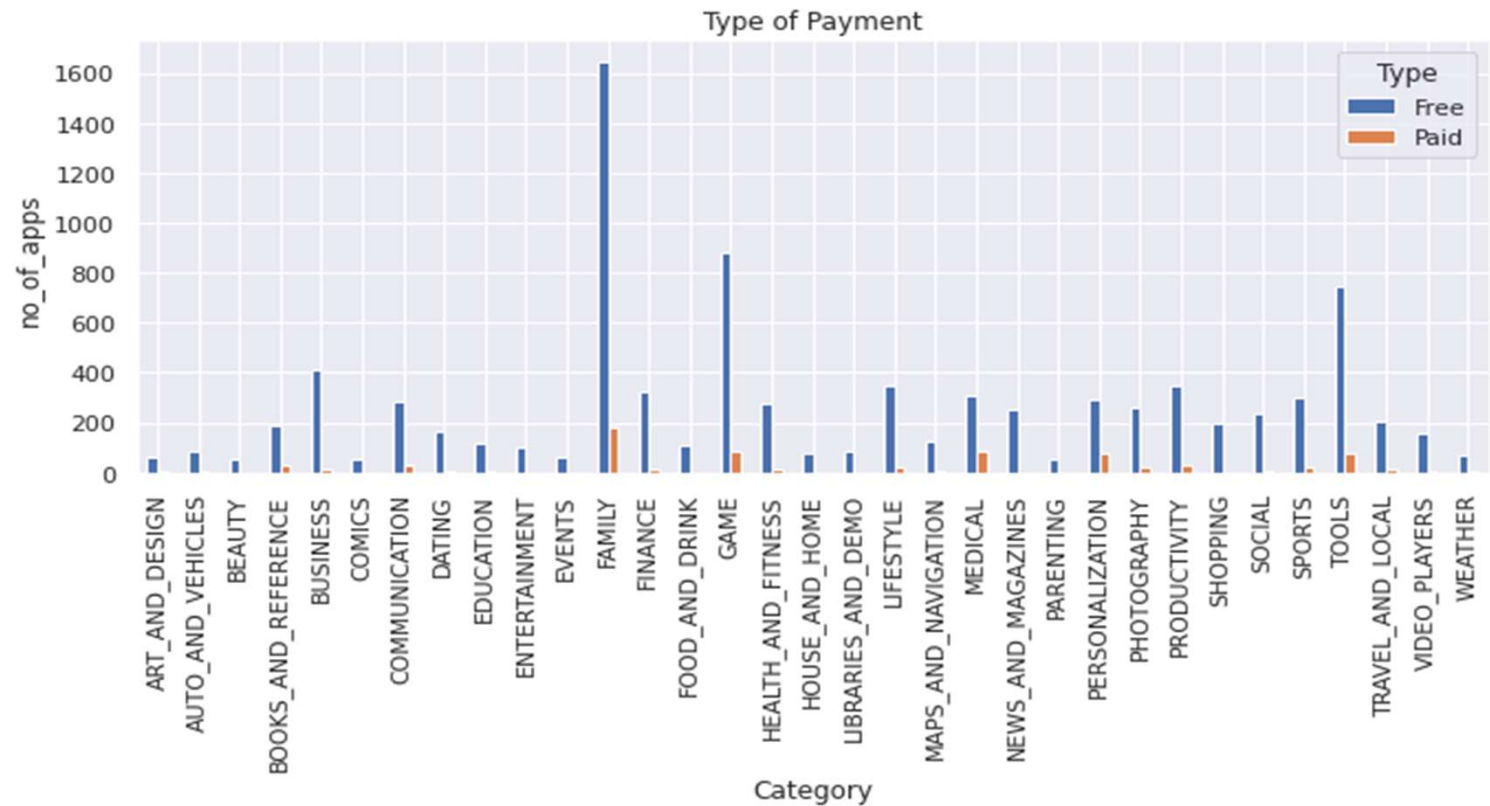
Maximum app have size between 0 to 30 , rest all are Outliers & Maximum Rating is between 4 to 4.5 rest all are outliers

Percent of Free Vs Paid Apps in play store



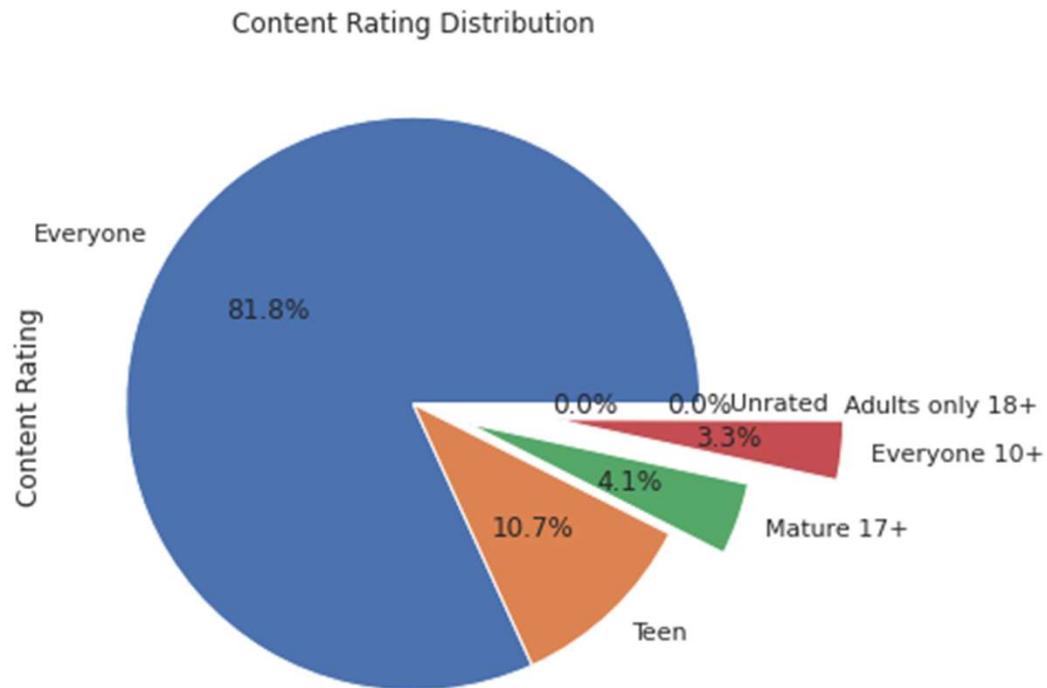
From this insight we can see play store have 92.2% Free Type and 7.8% is paid type

Number of Apps & its Category based on Type



Each category of free and paid types are shown here with number of apps (blue-free type and orange-paid type). It can be easily seen that there are lot more free apps as compared to the paid apps. Family category has highest free and paid apps

Based on Content Rating



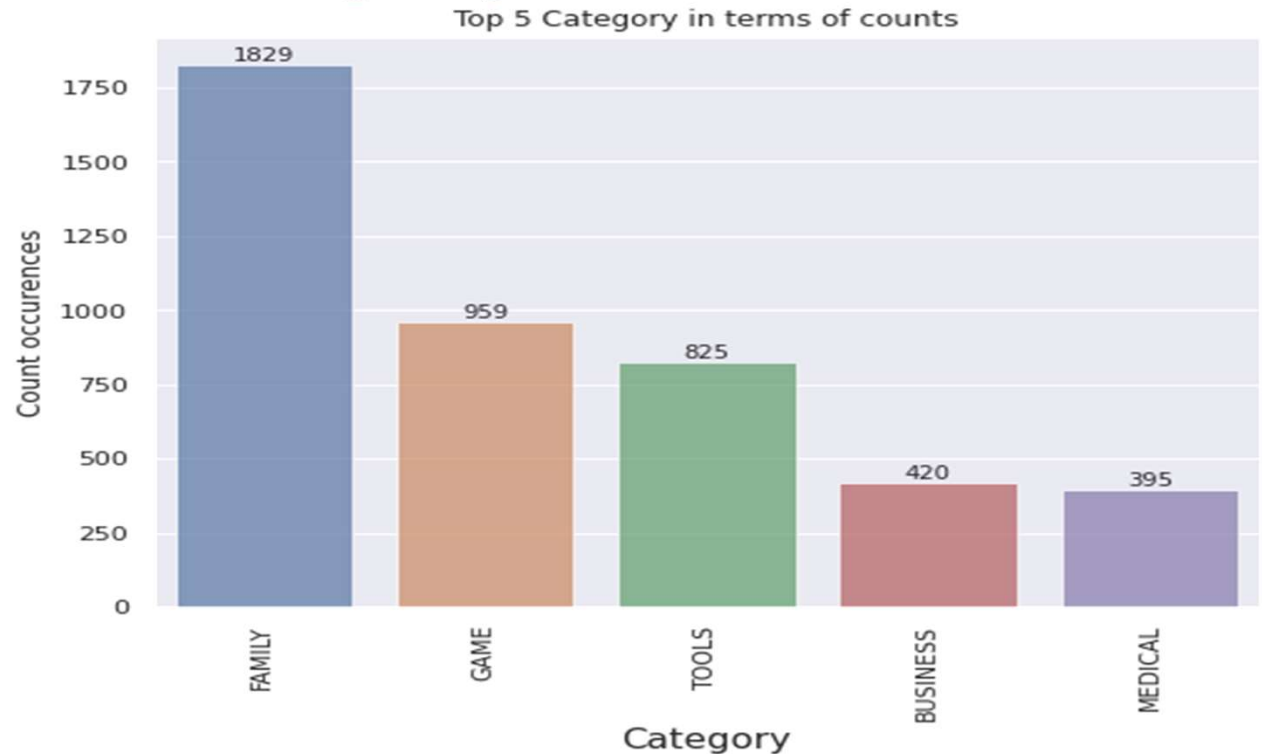
Everyone have greater portion in content rating Highest Rating Done by Everyone Then Teen Then 17+ Mature And 10+ Everyone Have almost Same Contribution and No Rating Done by Adults And unrated

Top 5 Based on Category



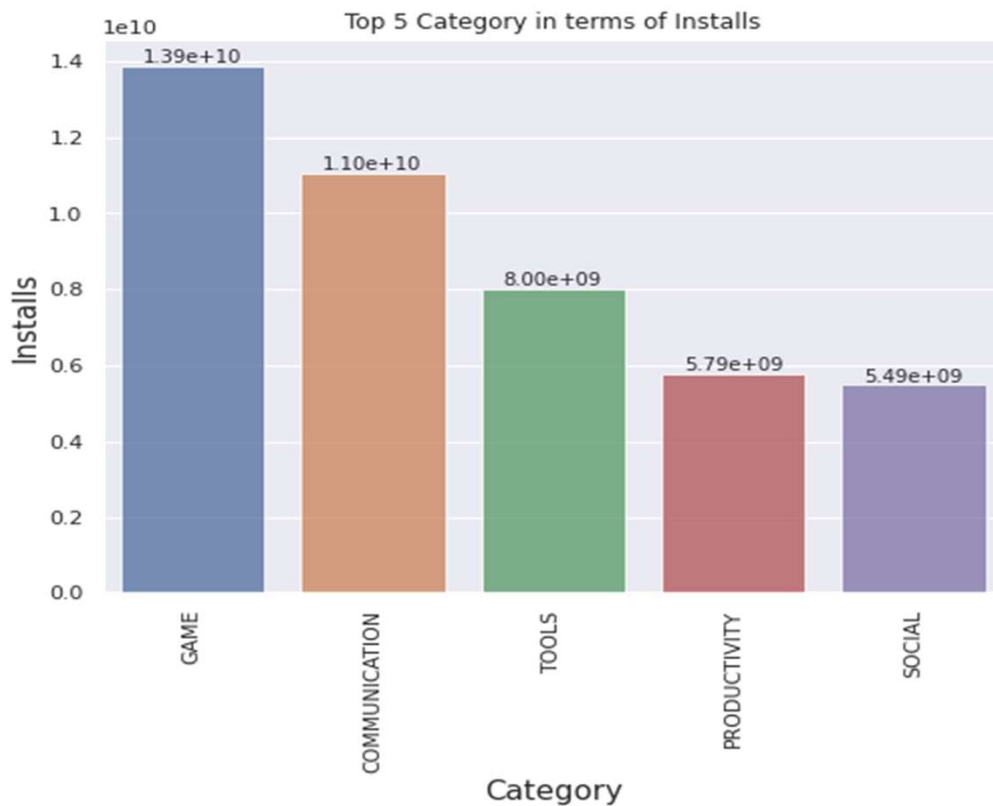
T O O L S

B



This bar graph is showing the analysis of top five category in terms of apps count. We observe that the category 'FAMILY' has the highest number of count, i.e, 1829.

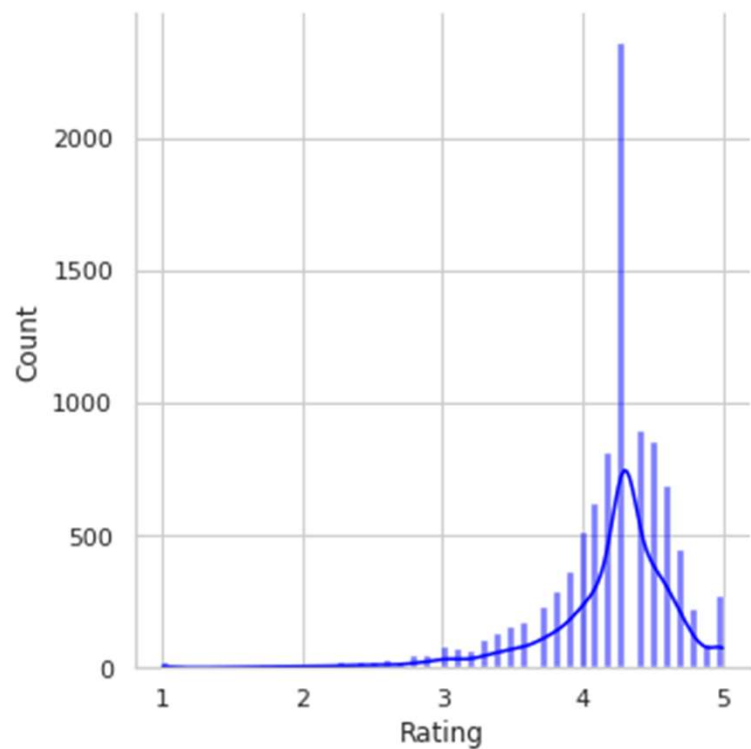
Top 5 category in terms of installs.



This bar graph shows the top five category in terms of installs. The category 'GAME' has the highest number of installs.



Based on Rating



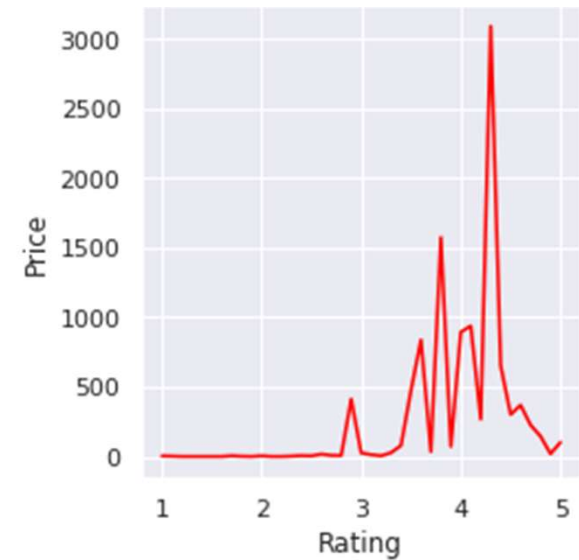
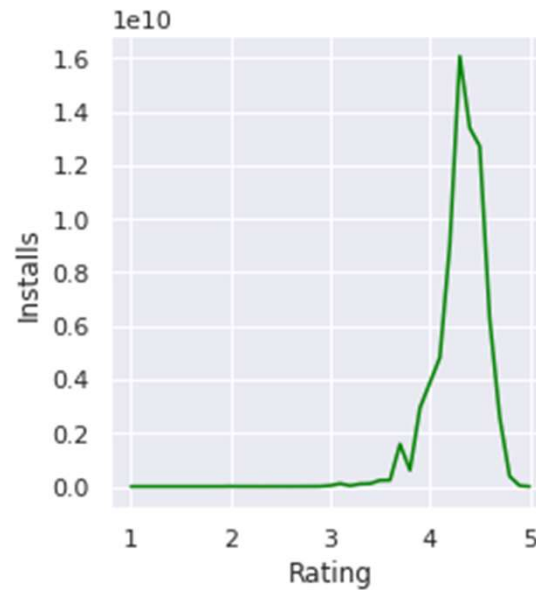
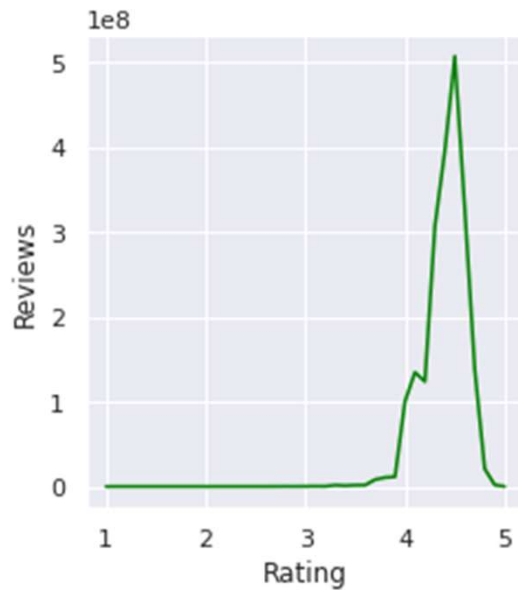
Highest Number of Rating is of 4.5



Rating



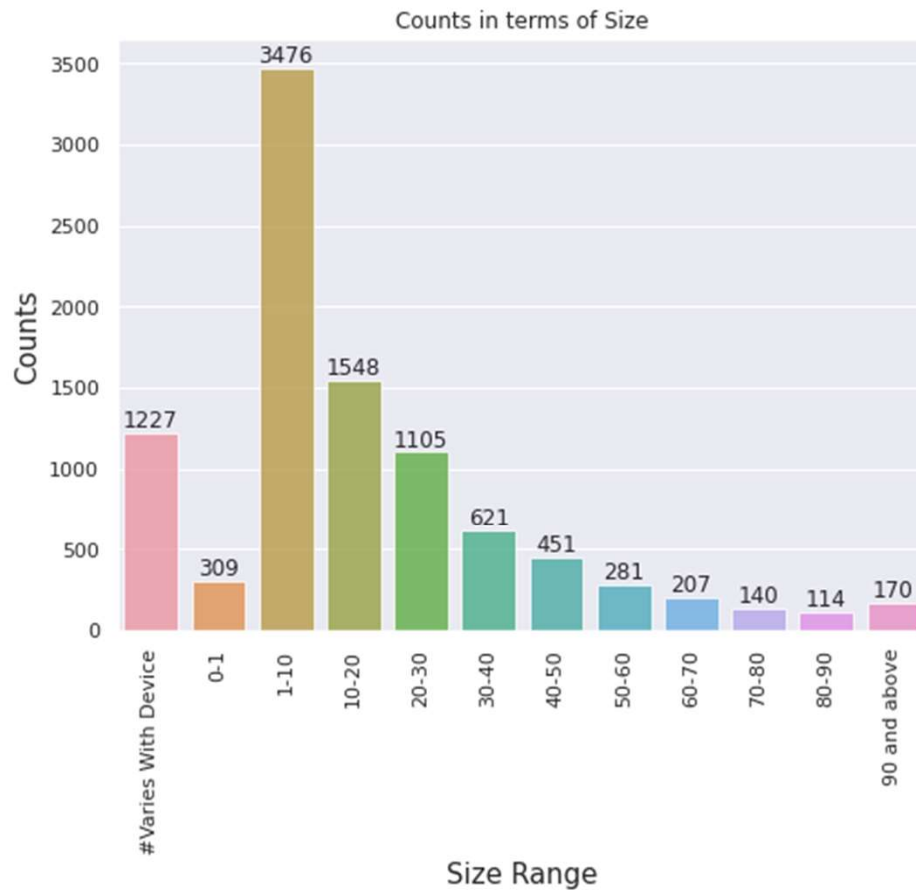
Rating



In analysis of rating pattern most of the app has rating between 4.0 to 4.5 w.r.t Reviews, Installs, Price



Based on Size Range



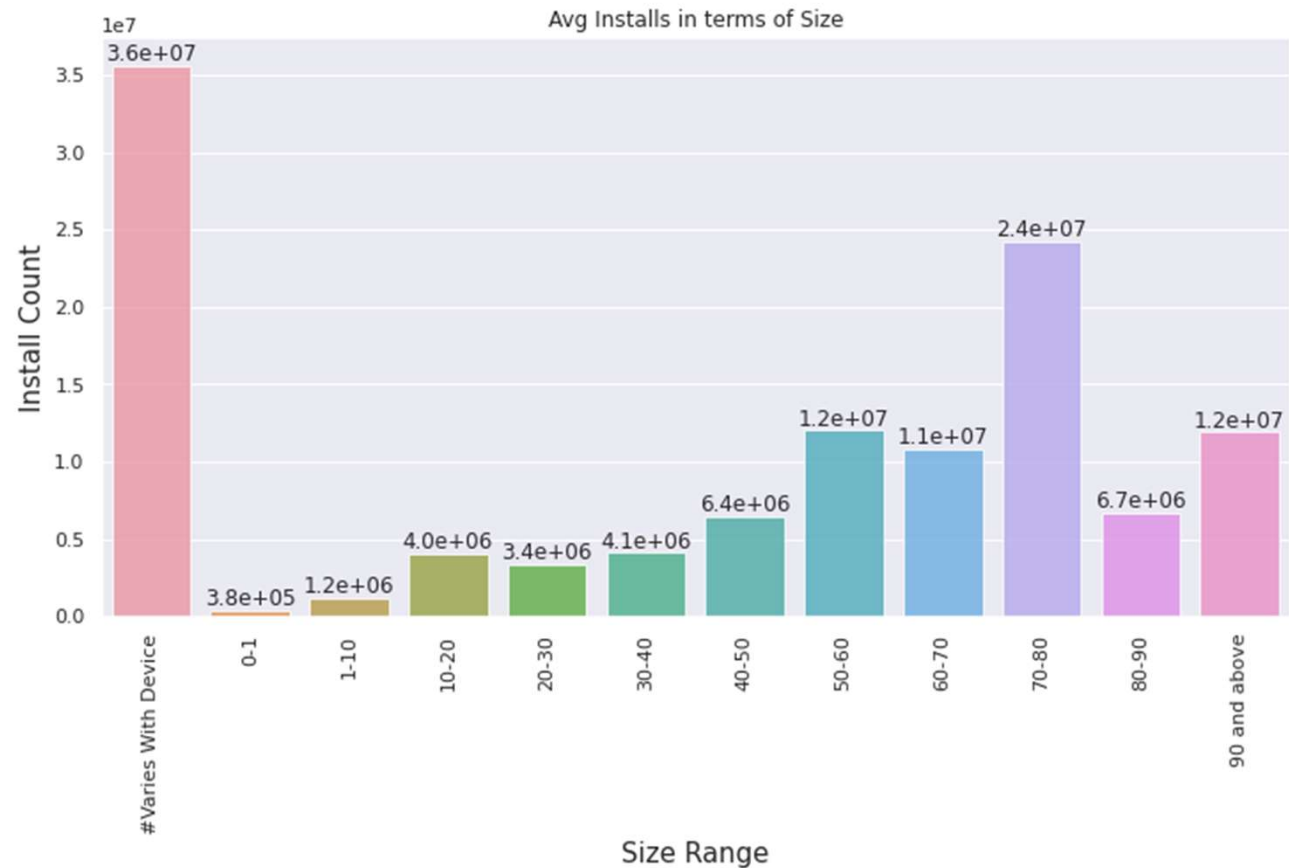
Most of the app has size between 1-10 & the fact is the smaller the size the higher the installation



Average Installs in term of size

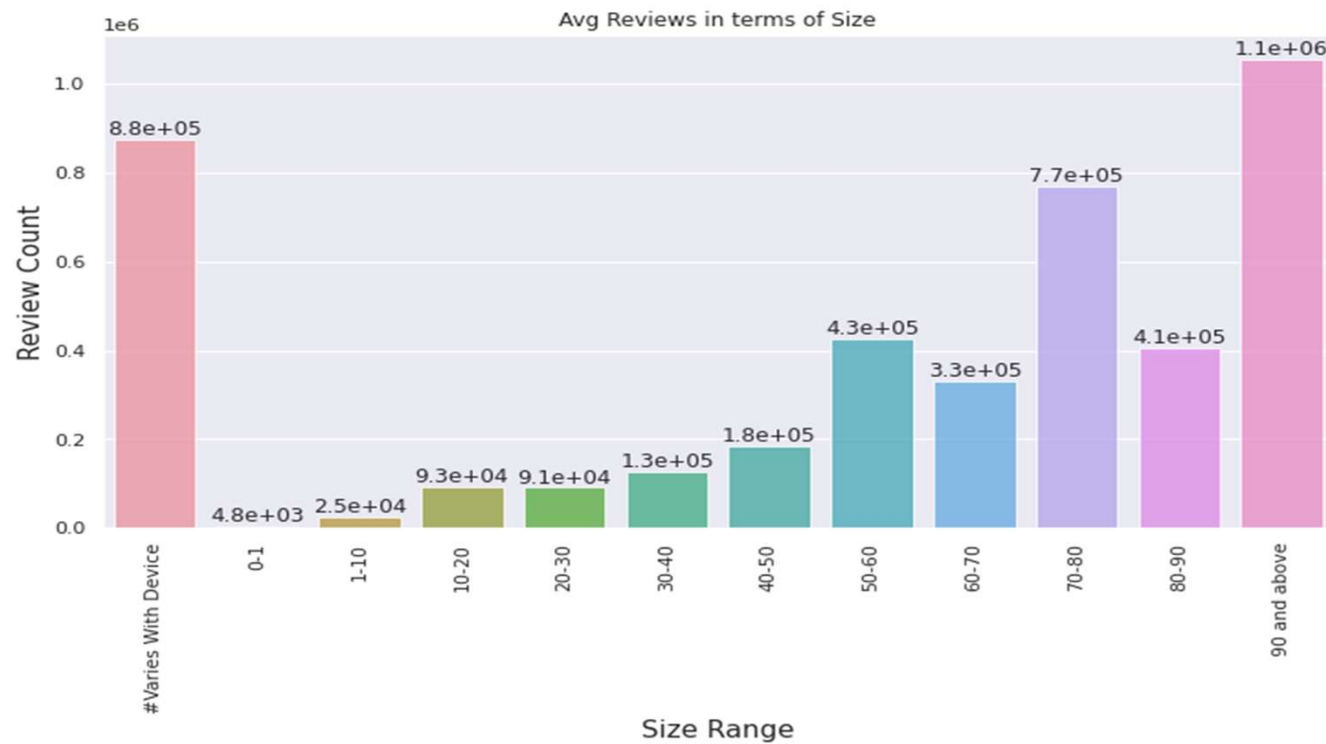


Average installs is above 3.5
Count and Size Varies from Device



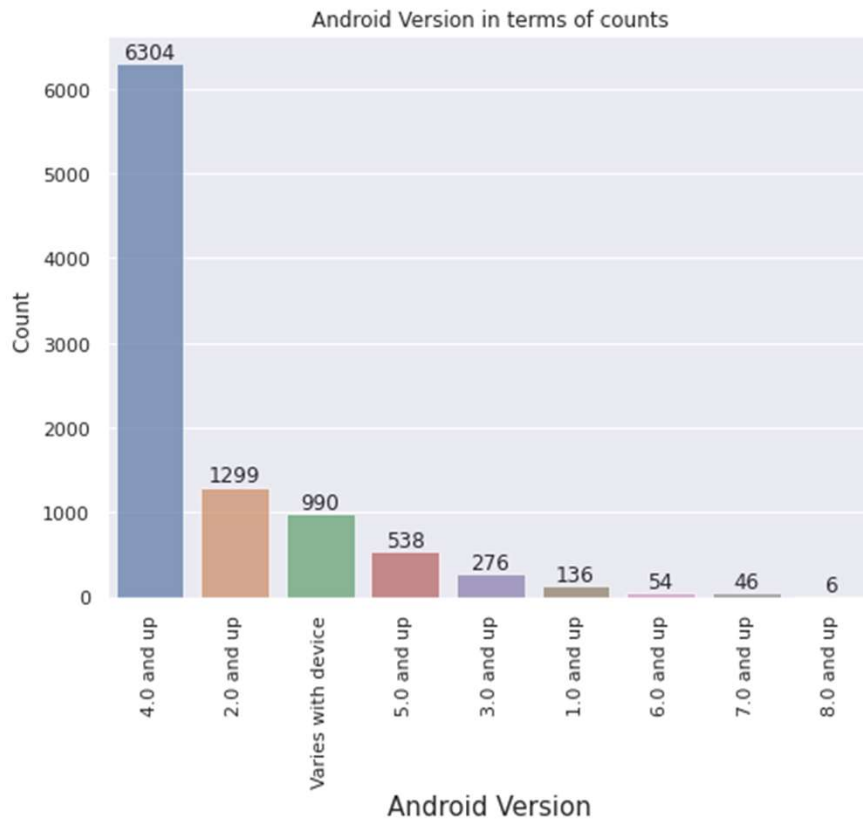


Average Review in term of Size



Average size of an app based on Review is 90 and Above

Based on Android Version



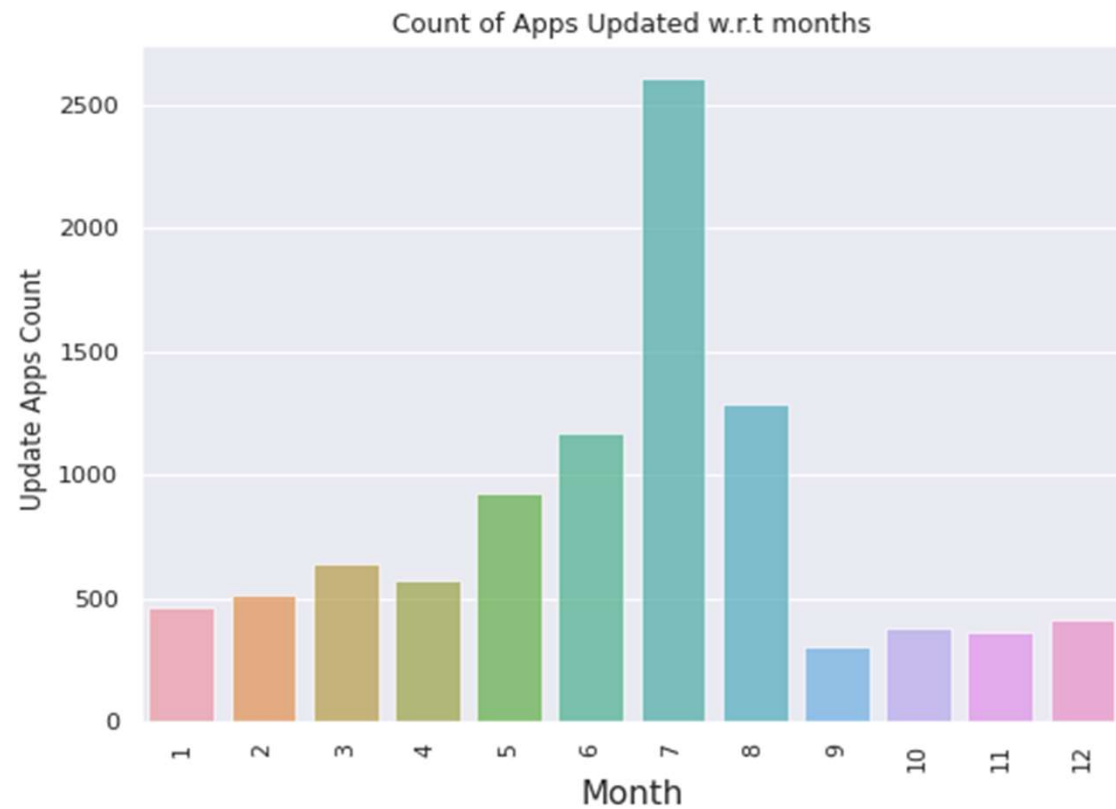
In analysis of android version then most of the apps are for version of 4.0 and up



Based On Last Update

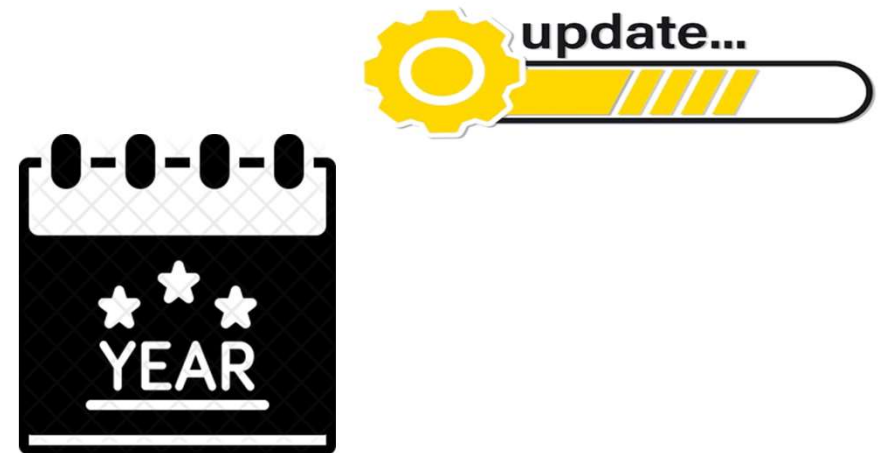
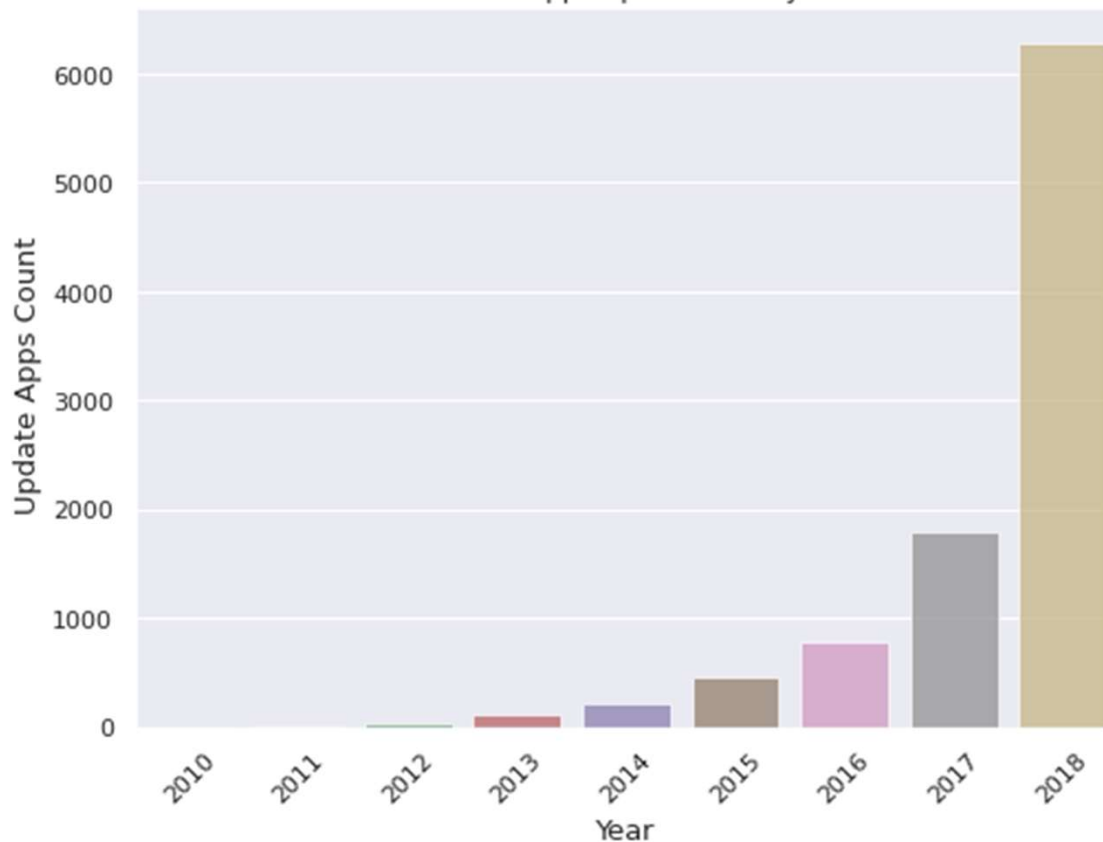


If we consider app update month then most of the app update in July



Updates With Years

Count of Apps Updated w.r.t year

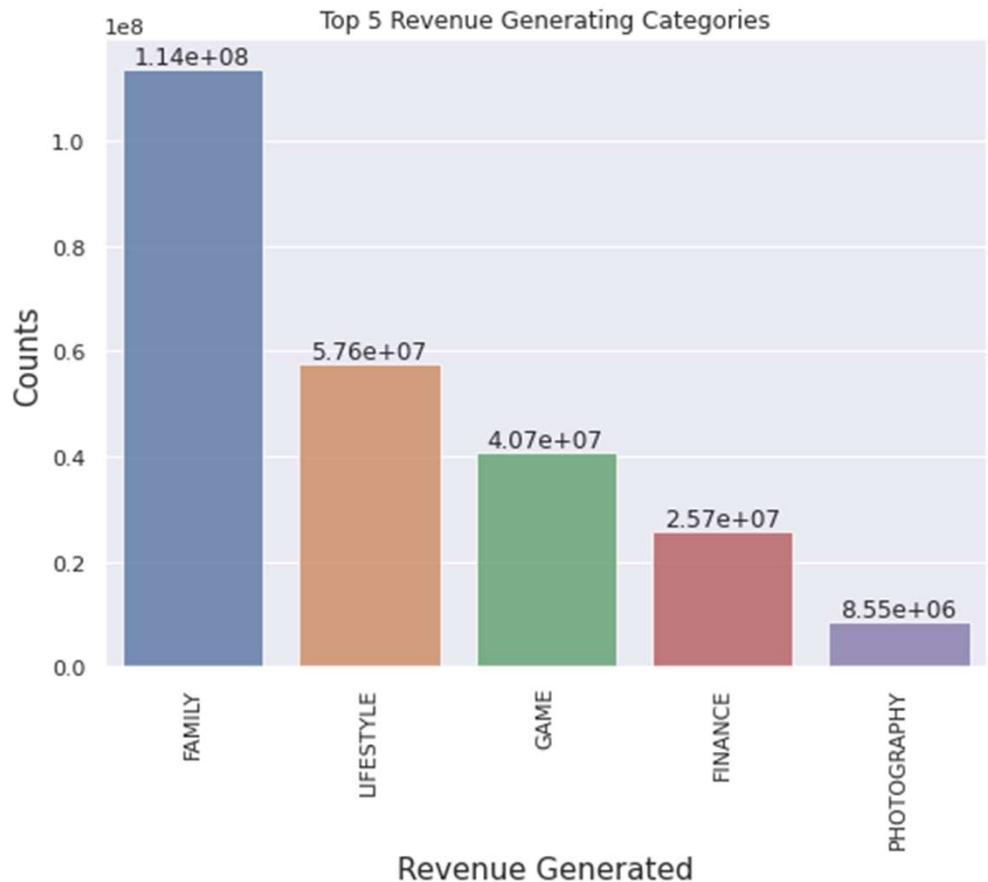


In analysis of app update year wise then updation of app increasing year on year basis . highest updated in 2018

Top 5 Revenue Categories



If we analyse the data in terms of revenue then Family category app has highest revenue



Conclusion

- ✓ In the free apps 64.1 % apps has positive sentiment followed by 22.4 % and 13.6 % for Negative and Neutral sentiment respectively
- ✓ In the paid apps 78.1 % apps has positive sentiment followed by 15.1 % and 6.7 % for Negative and Neutral sentiment respectively
- ✓ From the Sentiment , review graph we can see that no of installs 10000000 and 100000000 has highest review with positive sentiments
- ✓ Installs and review has a positive correlation
- ✓ total 92.2 % apps are free and 7.8 % apps are paid in play store data
- ✓ Family app category contains highest free and paid apps
- ✓ with content rating analysis 81.8 % apps are of Everyone content rating
- ✓ In analysis of top Five category in terms of apps count. Family has highest count i.e. 1829 apps and lowest Medical 395
- ✓ In analysis of top Five category in terms of Installs. Game has highest installs and lowest is social
- ✓ In analysis of rating pattern most of the app has rating between 4-4.5
- ✓ In analysis of app size pattern most of the app has size between 1-10
- ✓ In analysis of android version then most of the apps are for 4.0 and up
- ✓ If we consider app update month then most of the app update in july
- ✓ In analysis of app update year wise then updation of app increasing year on year basis . highest updated in 2018
- ✓ If we analyse the data in terms of revenue then Family category app has highest revenue



THANK YOU!